UNCLASSIFIED

Australian Government
Department of Defence
Defence Science and
Technology Organisation

A Feasibility Study of View-independent Gait Identification

Robert Whatmough

Intelligence, Surveillance and Reconnaissance Division
Defence Science and Technology Organisation

DSTO-TR-2684

ABSTRACT

This report reviews work on the identification of people in video recordings by their gait, or walking style. It considers the adjustments needed when the subjects walk other than straight across the image and may not offer the convenient side (fronto-parallel) view. It describes a study of this case using video sequences of walking people, and concludes that the available measurements of gait, apart from head height, are of limited use.

RELEASE LIMITATION

Approved for public release

UNCLASSIFIED

Published by

Intelligence, Surveillance and Reconnaissance Division
DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh South Australia 5111 Australia

Telephone: (08) 7389 5555
Fax: (08) 7389 6567

AR-015-265
March 2012

APPROVED FOR PUBLIC RELEASE

A Feasibility Study of View-independent Gait Identification

Executive Summary

The automatic analysis of gait, or walking style, of people in video imagery is of great interest as a possible means of matching them with people recorded in a database, at another time or by another nearby camera in a security CCTV network. Existing reviews indicate that gait has been studied in imagery for over twenty years. Gait has the advantage that some aspects of it can be studied in distant views where other features such as the face are not sufficiently resolved for identification.

Most human identification is an attempt to find the best match between a single person and many people in a database. In the security situation it is just as likely to be a comparison between two people who are otherwise new.

Gait is easiest to analyse when the person walks straight across the image and is viewed from the side. Different features of gait appear in a front or rear view, and all features may be present but in weaker form when viewed from intermediate angles. Viewing subjects from above head height adds useful information for the analysis of gait.

Other work on gait identification is reviewed, attention being paid to methods that allow for arbitrary points of view. Some of these methods are incomplete for gait analysis, but could form part of the process.

An experimental study using video sequences of people walking is described. The study has shown that the only gait characteristic that is very useful in distinguishing people and measurable from an above-head camera in any direction is the height.

Any further investigation of the use of other gait characteristics will require more extensive test data and more reliable ways of distinguishing walking persons from their backgrounds.

This page is intentionally blank

Author

Robert Whatmough
Intelligence, Surveillance and Reconnaissance Div.

Robert Whatmough was employed as Cadet Defence Science at the Weapons Research Establishement, Department of Supply, later DSTO. He was awarded the degree of B Sc with honours in 1969.

Until 1985 he was Experimental Officer, Scientific Officer then Research Scientist in Computing Services Group. His work included mathematical data handling techniques, random number generation, printed circuit board design, simulation of batch computer operations, curve fitting and smoothing, time-critical computing, thermal modelling, computer graphics and display, remote sensing and image processing, and assistance of computer users with complex computing problems.

Since 1986 he has worked in various Divisions in fields related to image processing. These have included restoration, enhancement and classification of visual, multispectral and synthetic aperture radar images, geographic information systems, prediction of oblique aerial and navigation radar images, matching models to objects in images, shape inference from perspective distortion and video sequences, registration and mosaicking of aerial images, and enhancement of video sequences.

This page is intentionally blank

Contents

1. Introduction

The automatic analysis of gait, or walking style, of people in video imagery is of great interest as a possible means of matching them with people recorded in a database, at another time or by another nearby camera in a security CCTV network.

Existing reviews (e.g. Gavrila, 1999; Moeslund & Granum, 2001; Wang *et al*, 2003) indicate that gait, as well as other forms of human movement and interaction, have been studied in imagery for over twenty years. The studies of gait have been inspired by reports of psychological experiments in which participants identified human movement by the motions of lights attached to selected points of people walking in the dark, or identified their friends by their gait alone. Gait also has the advantage that some aspects of it can be studied in distant views where other features such as the face are not sufficiently resolved for identification.

This report is concerned mainly with matching people between video sequences taken by different cameras at about the same time, the so-called "camera handover problem" (Redding *et al*, 2008). Gait can contribute to the evidence for or against this match. Gait is most easily analysed when people walk straight across the view, and the view is called "fronto-parallel". In realistic security monitoring situations, some people will walk in other directions, directly or obliquely, towards or away from the camera, and important features of their gaits will not be so easily detected, so the more general view needs to be considered.

In the rest of this report, Section 2 attempts to position the camera handover problem among the standard identification problems. Section 3 considers the difficulties that arise from less convenient camera viewpoints. Section 4 gives a partial review of other work in this area. Section 5 introduces a more view-independent approach. Section 6 considers the difficulties of locating a walking person accurately in a frame. Section 7 describes the view-independent method and a variation applicable to overhead views of shadows cast by walkers. Sections 8 and 9 report tests on real data. A final discussion and conclusions follow.

2. Camera handover as a matching problem

Human identification, whether done using gait or other characteristics, is most often a comparison of a new record (the "probe") with records of known people in a database (the "gallery"). The goal of comparison is classified as one of the following:

1. Identification: The person in the probe is known to be in the gallery. Which one is it? If there is uncertainty, which few could it be? The results will be judged by how often the correct person is picked (the "recognition rate"), how often it is among the few as a function of the number of suggestions allowed (called the "cumulative match score", CMS), the number of correct matches vs. the number of incorrect ones as a function of some parameter (the "receiver operating characteristic", ROC), or the pattern of correct and incorrect identifications (the "confusion matrix").
2. Validation or verification: The person in the probe claims to be one, possibly a particular one, in the gallery. Is that possible? The results can be judged by the ROC or the CMS.

3. Watching: The person in the probe is possibly one in the gallery. Is that so, and if so which one? The results can be judged by the ROC.

There may not be much difference between goals 2 and 3 in practice. Perhaps the comparison will be stricter for goal 2, but different prior probabilities of match and different costs of errors will affect the analysis.

In the case of camera handover, there is no gallery to start with. Rather there are two probes, and a person of interest in one is to be detected in the other if also present there. This goal resembles goals 2 and 3 above, with a gallery of one. Perhaps the timing and direction of movements at the two cameras will indicate which goal is more similar. When people of interest are completely unknown in advance, any collection of gallery data will be of use only in an experiment to test how much real and known people vary from one sequence to another and differ from person to person, and thence how strict the matching process should be.

3. The effects of changing view

The convenient horizontal and fronto-parallel view of a person walking before a stationary camera shows a number of features that can be measured and used for identification. Each foot in turn is raised from the ground, swings forward past the other and strikes the ground again, then remains more or less at rest for a time. After half the cycle (unless there is a limp) the silhouette repeats its shape sequence though the legs have swapped their roles. The length of the pace (“stride”) and the number of paces in a given time (“cadence”) can be measured and the subtler differences of leg motion are visible. Vertical motion of the head and forward or backward leaning of the torso are clear. The arms may swing in a characteristic fashion, though carrying objects such as bags or using a mobile phone can have profound effects.

When the walk is towards or away from the camera, most of the movement is in these same directions and may be difficult to detect at all in the silhouette or in the details of the clothing. The feet may not be lifted far enough from the ground for the vertical motion to be detected in a distant view; the rise and fall of the head and the swinging of the arms may remain more obvious. The shape sequence repeats after half a cycle but in mirror image. Progress may be evident only from a slow change in size of the figure, and stride may be hard to estimate. It is left to finer details such as the alternate exposures of the soles to the rear, changes in the wrinkle patterns of loose clothing and shadows cast by one leg on the other or between torso and arms to reveal the motion and its cadence. The roll of the hip (which rises on the side of a leg that is in contact with the ground, sometimes with an initial jerk as the heel strikes, and falls on the other side where the leg swings forward) will be much easier to detect at this orientation, but may require good resolution and contrasting upper and lower clothing near the waist (or a contrasting belt) to reveal it. Swinging the arms across the direction of walk instead of parallel to it will also be more obvious.

At intermediate angles, the characteristics visible from the side and those visible from the front or rear may all appear but in weaker form. The symmetry is gone, and the legs may cross over at times that are not equally spaced, or not at all. The cadence may still be clear but an estimate of stride may require one of the direction of walk.

Surveillance cameras are often installed above the head levels of passing people, so they face somewhat below the horizontal. This fact can help resolve some of the ambiguities of a horizontal view. People walking towards or away from the camera move respectively down or up the image, perhaps slowly. If suitable calibration is performed when the camera is set up, the mapping from image pixel to ground position is known and stride and direction can be estimated for any walk from the sequence of foot resting positions, once these are decided from the subject's motion.

4. Review of other work

This section reviews some of the work done on the various steps of gait analysis that is relevant when more general camera positions are to be allowed. The review is not complete. The other reviews mentioned in the Introduction can supply references to more work in this area.

4.1 Preparation of targets

Some researchers describe experiments where human subjects were marked at key points on their clothing, or asked to wear particular clothing, to simplify their detection and movement analysis. This is appropriate for medical diagnostic applications of gait analysis, but not for security applications.

4.2 Camera setup

Some authors (e.g. Knossow *et al*, 2006) describe experiments where several cameras were used simultaneously and their outputs were combined to give better 3-D constructions. Others (e.g. Wang *et al* , 2002) used several cameras quite separately to test how well their algorithms worked at different angles of view. Some (e.g. Kuno *et al*, 1996, Han *et al*, 2005) put records of the same subject walking in different directions in the gallery so that a match with any one view of a person was sufficient for identification.

4.3 Segmentation

The humans in a scene must be separated from the background, from their own shadows, from one another when several are present even perhaps in a crowd, and from other moving objects such as animals or vehicles. Many researchers have, however, assumed simple scenarios with a single moving human and a stationary background.

The usual approach to separating moving objects from a stationary background is to find an average over frames (often the median) and then to detect which pixels in which frames differ significantly in colour from the average for the same pixels (e.g. He & Debrunner, 2000). Sometimes a least-median-of-squares value is used instead of the median (Wang *et al* , 2002). Various forms of filtering can be used before estimating the background or after classifying pixels to improve the segmentation accuracy. Size constraints can be applied to distinguish humans from other objects, provided that distances to objects do not vary too much. The process will need to be adaptive if lighting conditions change over a long recording period.

If a human casts a shadow on the background, this can be absorbed into the background, using colour normalisation (e.g. Bobick & Johnson, 2001). Then the shadow is not counted as part of a foreground object despite its movement.

The human, once located, can be treated as a mere silhouette whose changing shape is to be analysed, or as a textured object with features to be retained for further matching. Many authors have reported useful gait identification by the silhouette alone, and most take this approach (Nixon *et al*, 2006).

4.4 Extraction of feature vectors

Gait matching methods differ mostly in how image sequences (binary or otherwise) are reduced to vectors that retain the essential features while removing superfluous ones and allowing fast comparisons. Most of them limit how much gait information can be extracted, especially when the view is not fronto-parallel.

Typically, the ranges of X and Y coordinates in the human segment (its “bounding box”) are found first. The bounding boxes in different frames are usually rescaled vertically to match and shifted horizontally and vertically to align, removing position and size differences. They may be rescaled horizontally as well to reduce shape differences. Possibilities for the next step include:

1. Using the bounding box contents as they are (Huang *et al* , 1998)
2. Using the bounding box contents for a few “key frames” (those showing the instants when legs were passing each other or both feet were on the ground) (Collins *et al*, 2002)
3. Summing the binary silhouettes over frames getting the “gait energy image” (Han *et al*, 2005)
4. Using only the lower (leg movement) part of the gait energy image (Bashir *et al*, 2008)
5. Taking the Discrete Fourier Transform of binary silhouettes over time and retaining the first few frequency coefficients for each pixel (Akihara *et al*, 2006)
6. Summing the binary silhouettes over rows and over columns separately giving two two-dimensional “frieze patterns” (Lee *et al*, 2007)
7. Dividing the bounding box into fixed-size parts and measuring the widths of the silhouette within them (Kuno *et al*, 1996)
8. Dividing the bounding box into fixed-size parts, using their corners or centres to represent body points and measuring the distances between them (Bobick & Johnson, 2001)
9. Representing the silhouette as a sequence of boundary points (Wang *et al*, 2002)
10. Combining the silhouettes into a space-time surface and finding the extreme points of its curvature (Yilmaz & Shah, 2008)
11. Fitting the silhouettes with a three-dimensional model and recording model parameters (e.g. Knossow *et al*, 2006; Bouchrika & Nixon, 2007; Brubaker *et al*, 2007) (Not all these authors completed the extraction of gait parameters from their fitted models.)

Further reduction of dimensionality can be done by Principal Component analysis (e.g. Bhanu, 2007), Linear Discriminant Analysis and Relevance Component Analysis (e.g. Tan *et al*, 2007), Local Linear Embedding (e.g. Li *et al*, 2005) or Latent Variable Methods (e.g. Cheng *et al*, 2008). These methods require a training phase using gallery entries to decide how many dimensions, and which ones, should be retained for the most reliable identification. The training does the work of deciding what it is that varies from one gait to another, something that may not be obvious to the human observer.

If the silhouettes are not resized and aligned first, summing them over time produces an image in which the largest values occur at the places where the feet rested on the ground. If these can be related to ground positions, they yield a measure of stride, and also cadence if the timings of passages are retained. The procedure can be refined by counting only the pixels in each frame near corners of the silhouette edge in that frame (e.g. Bouchrika & Nixon, 2007).

4.5 Cycle detection

The gait cycle (of two paces) needs to be distinguished at some stage, either before feature vector extraction (so that averages over time are taken over an exact number of cycles) or after it (so that measurements can be made at the relevant times to estimate the cadence and other parameters). The distinction must be made using preliminary measurements of foot position or separation, the minimum and maximum widths of the silhouette at foot level, or some other measure of similarity between frames.

4.6 Incorporating other data

Some authors considered the use of other data such as face appearance as well as gait (e.g., Zhou & Bhanu; 2008, Shan *et al*, 2007; Huang *et al*, 1998). While simple combination rules such as "A mismatch of any type is a mismatch overall" are useful, combination was usually done by concatenating feature vectors or by combining principal component coefficients using a more sophisticated method such as Canonical Correlation Analysis.

4.7 Allowing for different view points

Most authors assumed that the view was fronto-parallel in their analyses, though some included different view points in the gallery, to be matched in different cases. One paper measured the rapidly declining performance of two methods when the view point in the probe moved away from the fronto-parallel direction used in the gallery (Bashir *et al*, 2008).

Li *et al* (2005) claimed that when they reduced silhouettes to one dimension (using Local Linear Embedding, LLE), the single coordinate related to body extent in a view-independent way. They further claimed that a higher-dimensional LLE representation is independent of translation, rotation and scale, so it was a good basis for further work on view independence.

Yilmaz & Shah (2008) claimed that their process of constructing a surface in space-time from all silhouettes had a view-point dependence that was eliminated by later steps of their method. They proved this for the case where a camera sees the same points of a three-dimensional object from each view point, but failed to note that cameras at different positions see different points of an object's surface as points on its silhouette. Thus their claim of view-point independence fails.

Bobick & Johnson (2001), having chosen body points whose identification is not very sensitive to the view point, used some of their measurements between them to correct others and reduce the effect of changing view point.

Rao & Shah (2001), who were trying to recognise hand actions rather than gaits, tracked the centroid of the fastest moving region of skin colour, identified points of maximum curvature in its space-time path and formed lists of signs of direction changes at these points. If two lists matched, they constructed a matrix from the trajectories and checked its rank, in a manner related to the factorisation method of structure-from-motion reconstruction (e.g. Tomasi & Kanade, 1992), to confirm the match. They argued that the list of signs was not affected by most changes of camera setup. Parts of their approach could be useful once particular points on a walking person, especially hip joints, knees and feet, are identified.

Tresarden & Reid (2008), who were dealing with other body actions as well as walking, used a structure-from-motion factorisation algorithm that depended on symmetry constraints (such as legs of equal length) rather than rigidity to locate joints in space. They depended on markers on joints and did not go on to identify the actions, but their approach might work on features detected from silhouettes.

Spencer and Carter (2005) used another form of shape-from-motion reconstruction in which periodicity of gait, constancy of limb lengths and movement of limbs in parallel planes are all assumed. The assumptions are not always correct (arms sometimes being swung at an angle, for example), but the method suggests another variation. If walking is truly periodic, frames selected at the same point in many cycles should show an apparently rigid body in linear motion past the camera. If the person is not distant, these frames might be used to perform a simple structure-from-motion reconstruction by factorisation. The same process repeated at other points in the same cycles would provide one cycle's worth of structures. After suitable alignment, these too could allow temporal analysis of the gait. Allowance might need to be made for timing fluctuations using "time warping" (e.g. Kaziska & Srivastava, 2006).

A fairly crude approach to view invariance is to assume that the human body is two-dimensional, lying in its sagittal plane. If the actual direction of walk can be estimated, the change in silhouette can be adjusted using an affine transformation. Kale *et al* (2003) took this approach, using the overall direction of motion and camera calibration to get the walk direction.

Ben Abdelkader *et al* (2002) set up their camera at an elevated position and used knowledge of camera calibration to relate image coordinates to ground positions. They located the feet when they rested on the ground and deduced enough of the body position to determine stride, cadence, and the mean and amplitude of fluctuation of head height.

Akihara *et al* (2006) trained their method with different directions of walking to find a transformation of their feature vector that allowed for differences in view point. This transformation changed the view point to either fronto-parallel or front-rear, whichever was nearer, these two views then having to be treated separately.

Any of the methods described in Section 4.4 that fit three-dimensional models could form the basis of a view-independent gait identification system. The fitted model moves, and its motion relative to its own orientation must be reduced to some suitable vector of gait parameters for matching with models fitted to other sequences.

5. Selection of a view-independent approach

The review above, while incomplete, suggests that identifying people by their gait from near-horizontal but otherwise arbitrary directions is not a mature technique. Such identification has however been done with some success.

Attempting to match body features created by clothing appears to be unreliable, at least for uncooperative subjects, because too many features are merely transient creases or shadows. Depending on these features reduces the usefulness of gait for identification at a distance. Without them, only the outline of the body is available for identifying feature points, so they may be limited to head, feet and main joints. Even then, loose clothing such as a skirt, robe or overcoat, and concurrent activities such as carrying a bag, will further limit what can be identified.

Sections 3 and 4.4 above noted that it is possible to determine, from a video sequence from a camera directed obliquely from a position above head height, when and where the feet rested in the view and thence where they rested on the ground. For a suitably calibrated camera, if an object in the air is known to be vertically above another on the ground, its height can be determined too. There should then be enough information to decide details of stride and head height, just from the sequence of silhouettes.

The next sections describe a method for gait analysis by silhouette extraction, partly inspired by the contribution of Ben Abdelkader *et al* (2002).

6. Silhouette extraction

It will be assumed that a walker is detected and reduced to a silhouette before any gait analysis is performed. In this way, features of clothing take no part in the identification of body parts. Although the extraction of silhouettes is not the main subject of this report, some was needed in the early preparation of test data, and techniques were developed that could be useful in other experiments or security applications.

6.1 Raw detection

The usual approach to detecting moving objects is to identify the background (assuming that any part of it is visible in most of the frames in the sequence) and then to detect which pixels in which frames differ from their background values and so belong to the foreground. The latter step is known as "background subtraction". The use of colour is essential here. Identifying the background and keeping it up to date are non-trivial (especially outdoors) and will be taken as already done.

The complications of identifying foreground pixels include:

1. The presence of irrelevant moving objects such as clouds, vehicles, animals, vegetation that is disturbed by the wind and the lighting changes caused by all of these
2. Shadows cast on the background by walking people
3. Highlights caused by extra illumination of parts of the background by walking people
4. Accidental resemblances of parts of people (including their clothing) to parts of the background

The normalisation of colour images (so that all pixels have the same luminance, an appropriate weighted sum of the colour components) can greatly reduce the effects of shadows and other lighting changes, but it also blinds the detection process to major differences between the luminances of foreground objects and the background.

Elgammal *et al* (2002) described a method of background identification and foreground detection in which allowance is made for weak shadows but stronger luminance changes are still recognised as differences. The colour components (R, G, B) are transformed to two chromaticity variables and a lightness variable using

$$\begin{aligned} r &= R/(R+G+B) \\ g &= G/(R+G+B). \\ s &= (R+G+B)/3 \end{aligned}$$

If the transformed background values are (r_b, g_b, s_b), a pixel matches the background if (r, g) is close to (r_b, g_b) and $\alpha \le s/s_b \le \beta$, where $[\alpha, \beta] \ni 1$ is the range of lightness changes likely to be caused by shadows and highlights. Details of the definition of "closeness" here are submerged in the kernel density method, the main subject of the paper, but the approach suggests a "soft-thresholding" method.

Let the lightness s be defined as above, or as some other suitably weighted mean, and let s_b be its background value. Define the soft-thresholded value

$$s' = \left\{ \begin{array}{ll} s + (1-\alpha)s_b, & s \le \alpha s_b \\ s_b, & \alpha s_b \le s \le \beta s_b \\ s + (1-\beta)s_b, & \beta s_b \le s \end{array} \right\}.$$

Equivalently,

$$s' = s - \tfrac{1}{2}\left|s - \alpha s_b\right| + \tfrac{1}{2}\left|s - \beta s_b\right| + \left(1 - \frac{\alpha+\beta}{2}\right) s_b \,.$$

The modified pixel colour

$$\left(R', G', B'\right) = \frac{s'}{s}\left(R, G, B\right)$$

matches the background colour if their brightness values agree within the allowed range of shadow and highlight and their chromaticity values match too. Larger brightness differences and any chromaticity differences are retained, and a distance measure between modified and background colour components,

$$d = \left|R' - R_b\right| + \left|G' - G_b\right| + \left|B' - B_b\right|,$$

can be compared to another threshold to decide whether the pixel is in the foreground.

6.2 An orthogonal least-squares method

If there are features in the background that can be recognised within a shadow or highlight, it may be possible to avoid treating strong shadows or highlights as foreground.
This requires that part of the image within a window be compared to the same part of the background image, and that a correlation is detected between two sets of values.

Suppose that values x_i in the background are compared with values y_i for the same pixels in the current image. If both sets of values are affected by similar random errors of measurement, and the values are to be fitted by a straight line $x = u\cos\theta, y = u\sin\theta$, it is appropriate to minimise the sum P of squared perpendicular distances to this line. With allowance for different weighting of points in the sum, this takes the form

$$\begin{aligned} P &= \sum_i w_i\left(-x_i \sin\theta + y_i \cos\theta\right)^2 \\ &= \frac{1}{2}\sum_i w_i\left(x_i^2 + y_i^2\right) - \frac{1}{2}\sum_i w_i\left(x_i^2 - y_i^2\right)\cos 2\theta - \frac{1}{2}\sum_i 2w_i x_i y_i \sin 2\theta \\ &= \frac{1}{2}S_0 - \frac{1}{2}S_1\cos 2\theta - \frac{1}{2}S_2 \sin 2\theta, \text{ say} \\ &= \frac{1}{2}S_0 - \frac{1}{2}D\cos 2\left(\theta - \theta_0\right) \end{aligned} \tag{1}$$

where

$$D = \sqrt{S_1^2 + S_2^2}$$
$$2\theta_0 = a\tan 2(S_2, S_1)$$

P is minimised at $\theta = \theta_0 + n\pi$, which can be reduced to the range $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Negative θ_0 values indicate negative correlation, while values greatly exceeding $\frac{\pi}{4}$ indicate excessive highlighting.

If the luminance ratio is constrained to the range $\left(\frac{1}{T}, T\right)$ and $\tan\theta$ falls outside this range, P is minimised by setting $\tan\theta$ to the nearer end of the range and taking $\cos 2\theta = \frac{T - 1/T}{T + 1/T}, \sin 2\theta = \frac{2}{T + 1/T}\operatorname{sgn}(S_2)$ in equation (1). Otherwise the minimum value is $\frac{1}{2}(S_0 - D)$. An excessive value of P indicates that the fit is poor and that the image values are not compatible with the background even under shadow or highlight conditions.

This method can be efficiently implemented for images. The sums in equation (1) are done around all pixel positions by convolution. Colour is handled by treating the three bands as separate regions and combining their P values. A 5x5 box filter, and a threshold of $(0.05)^2$ for P, have been found partly effective for image values in the range $[0,1]$, but foreground detection remains error-prone, especially where a plain object lies in front of a plain background.

6.3 Refinement

If there are false background objects or "holes" in the true foreground, a useful technique is to take the raw silhouette frame, with pixel values say '0' for background and '1' for foreground, and find the 4-neighbour or 8-neighbour connected regions of foreground and background. Small regions of foreground can be changed to background to eliminate the clutter, and the remaining regions re-identified. Small regions of background can then be changed to foreground to fill holes.

In the simplest case, only the largest foreground region is retained, then small background regions are removed. If there are two or more large objects present, a more sophisticated analysis is needed.

7. The gait analysis method

7.1 Background of the method

The method of Ben Abdelkadar *et al* (2002) located the feet when they rested on the ground by looking for the gap between the legs, then deduced enough of the body position to determine stride, cadence, and the mean and amplitude of fluctuation of head height. This technique worked for fronto-parallel views (where the walk was perpendicular to the line of sight) and for diagonal views, but for walks along the line of sight there may be little or no gap. A more general approach is needed for detecting where the feet land.

When the ground is not moving and the camera is fixed and right way up, the most reliable features of feet are that they are near the bottom of the silhouette, that they have downward facing edges, and that they do not move when they are on the ground. It will be assumed also that they are visible most of the time (though one leg can sometimes occlude the other).

7.2 Footprint detection

The following method is proposed for locating the footprints (resting positions of the feet) in image space and in time:

1. In each frame, find the highest and lowest rows, and the leftmost and rightmost columns occupied by the silhouette. These form the "bounding box" of the silhouette.
2. In each frame, apply an edge detector to the pixels in the bottom part of the bounding box (the actual fraction to be determined by experiment), retaining only positive values of the upward component found by convolution with the kernel $\begin{bmatrix} -1/4 & -1/2 & -1/4 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$. Generate zeros elsewhere. This detects pixels on the lower edge of the lower part of the silhouette, giving less weight where the edge is not horizontal.
3. Sum the results over frames three times, once without weights, once with weights equal to frame numbers, and once with weights equal to frame numbers squared. This produces three images which record the sum of, and the first and second moments of frame number over, the edge values of each pixel.
4. Smooth the three images using a Gaussian or similar low-pass filter.
5. For pixels where the sum of edge values exceeds a suitable threshold, assume that a footprint is present and compute the mean and standard deviation of frame number from the sum and two moments. Then there are measures of when and for how long the footprint was occupied, for each relevant pixel.
6. Retain the highest row number for each frame as a record of head height for later use. Record also the horizontal position of the top of the head, as it may have a subtle effect on height at a later stage.

The above method can be applied in a single pass through the image sequence without retaining the contents of individual frames. (If there is any clutter in the silhouette frames, despite the attempt in step 4 to reduce it, it may produce false edges at inappropriate times and distort the results. If the outliers are to be detected and removed, there may need to be further passes through the sequence.)

There is nothing so far to distinguish walking from, say, sliding slowly across the view on ice skates. For walking, the footprint records for single pixels form clusters that are well separated in space and time. (Any overlap of contact times as the walker's weight is transferred from one foot to the other is eliminated by using the mean frame number for each contact after step 5.) The clusters can be found as follows:

1. Form a histogram of the mean frame numbers of all pixels where a footprint is present, with one bin for each frame. (An alternative would be to regress the row and column numbers on the mean frame numbers for these pixels and use the result to compute a "time of passage" for each pixel. The histogram would then be of the times of passage. This

approach was found to be too sensitive to errors in silhouette extraction, to changes in speed and direction of the walk, and to perspective effects.)

2. Smooth the histogram using a combinatorial filter. With an appropriate degree of smoothing, the local maxima will mark footprint times and the local minima will represent threshold times between footprints. (See Figure 4.)
3. For each footprint, delimited by thresholds from the previous step, assign to it pixels whose mean frame number is in the correct range. Combine their statistics of occupation to get a centroid in space and the mean and standard deviation of the frame numbers.

The walk is now represented by a series of footprints, each with an image position and a frame number. These can be analysed to determine gait characteristics. (The centroid of the foot is in fact the centroid of the nearer parts of the edge of the foot, the further parts being occluded by the leg or rejected by the edge detection criteria. For the present application, this bias is not very important.)

7.3 Gait features

From the footprint information, some useful gait features can be extracted. Initially these will be measured in image coordinates and frame numbers; camera information will be used later. The features are determined as follows (See Figure 1):

1. Half the displacement from a footprint to the footprint two paces later is the stride, represented as a pixel coordinate vector.
2. Half the difference between the mean times of footprints two paces apart represents the stride time as a number of frames.
3. The mean position of footprints one pace apart represents the approximate body position (as a point on the ground) at "double-stance", when both feet are on the ground. The mean position of two double-stance positions one pace apart represents the approximate body position at "mid-stance", as one leg swings past the other.
4. The displacement from the body at mid-stance to the print of the foot on the ground, classified as being to the left or right of the stride vector, identifies the foot on the ground and hence the left or right identity of every footprint. Its magnitude is half the leg separation (the distance between the line of left footprints and the line of right footprints).
5. The body position on the ground at double-stance can be compared with the head height recorded from the frame at the mean time of the two consecutive footsteps. The height difference in image rows represents the head height above the ground when it is least.
6. The body position on the ground at mid-stance can be compared with the head height recorded from the frame at the mean time of the one occupied footstep. The height difference in image rows represents the head height above the ground when it is greatest.
7. If the distances stepped by the left and right feet are measured separately and parallel to the vector of stride, their difference, as a fraction of their sum, is a signed value that represents the degree to which one foot is preferred to the other, that is, of limp.
8. The standard deviation of the frame numbers for which a footprint is occupied is approximately $1/\sqrt{12}$ of the time for which it is continuously occupied. (The estimate will be reduced slightly because the heel lands or lifts off before the toe does the same thing.) The times for which each footstep is occupied are therefore known as numbers of frames. If

the times are measured for the left and right feet separately, then their difference, as a fraction of their sum, is another measure of limp. (The signs of the two measures can be expected to agree.)

9. The sum of the times for which footprints are occupied should be a little greater than the total time for an image sequence. The difference, as a fraction of the total time, is a measure of the tendency to reduce the contact time of each foot with the ground, a feature of running (or at least hurrying).

For most of the above features, the estimate (or part of it) can be averaged over as many pairs of footprints as are available from the image sequence.

Camera information is needed before gait features can be expressed in time and length units. In the simplest case, the view is a distant one and perspective effects can be ignored. Then only three quantities may be needed: the width of a pixel on the ground in length units, the depression angle of the camera, and the frame rate (frames per second) of the camera. From these, the height of a pixel is known, both as a distance on the ground away from the camera and as a distance along a vertical line near the walker.

In the present work, perspective and radial lens distortions were found important. The lens distortion must be estimated by examining straight line features in the images, and corrected for in every measured image coordinate. (Image warping is not required.) For the camera setup, the quantities needed are the depression angle, the distance to the ground at the image centre, the focal length relative to the pixel size in the sensor plane and the rotation of the camera about its axis ("swing"). There are other sets of quantities equivalent to this set. The size of a pixel then depends on the location in a frame.

The following features are now available in world units or dimensionless form:

1. The stride in length units, as a vector or a length
2. The cadence, or number of steps a minute
3. The speed in preferred units
4. The leg separation
5. The limp indication from lengths of steps
6. The limp indication from lengths of stance
7. The run (or hurry) indication
8. The minimum head height at double-stance in length units
9. The maximum head height at mid-stance in length units

7.4 Gait analysis of shadows

Stoica (2008) recently considered the feasibility of using high angle views from aircraft or satellites to study gait from the shadows of walkers on the ground, early or late in the day. From such angles the view of the body of a walker is not suitable.

Assuming that sufficient resolution is available, the shadows contain the same shape information as a low elevation view from the direction of the sun. If camera position, orientation and time are known, gait information can be extracted by the following steps:

1. Register the frames so that the ground is stationary.
2. Extract the silhouettes so that the shadows are detected, rather than the bodies.
3. Correct for camera tilt to give a vertical view of the ground, if this hasn't been done as part of step 1.
4. Rotate the frames so that the sun is casting the shadows in the upward direction.
5. Calibrate the view by finding the size of a pixel on the ground (the same in the along-shadow and cross-shadow directions) and the height of an object that casts a shadow one pixel long (from the sun elevation).
6. Find the gait parameters in terms of image coordinates as before and convert to stride lengths and body heights using the modified calibration of step 5.

For some camera angles there may be problems, for legs can effectively occlude the shadows of the feet. It might be better then to include the body in the silhouette.

8. Tests on treadmill sequences

8.1 Methods

Initial tests were made using the MoBo database (Gross & Shi, 2001). This database provides many colour sequences of different subjects walking on a treadmill in several styles, viewed from several directions (Figure 2). The directions include walking directly and obliquely towards and away from the camera and straight across the picture. There is a single background frame for each sequence, and this is needed for parts of the background that are never visible during a walking sequence.

Silhouettes are given for all sequences, but these sometimes include walkers' shadows and effects of clutter. The techniques of Section 6.1 were tried and parameter settings were found that produced acceptable (but still imperfect) silhouette sequences for the five directions of walk just described (Figure 2).

On a treadmill, a foot at rest still moves and produces no "footprint" in the sense of Section 3. It was therefore necessary to modify the sequences so that they appeared to be on a stationary surface. This was done by visual inspecting each sequence to find sub-sequences where a foot was resting, determining the velocity in image coordinates of the treadmill mat during the sub-sequences, then inserting frames from the sequence into larger images at different positions chosen to reduce the velocity of the resting feet to zero. The results were then sets of sequences of the same subject walking with identical gait in different directions.

Figure 3 shows the sums of edge values formed during footprint extraction as in Section 3.2. Figure 4 shows the histogram of the mean frame numbers of footprint pixels, before and after smoothing.

Camera calibration was performed by locating in one frame of each sequence the corners of a floor tile near the treadmill and assuming that it was an orthographic projection of a 20 cm square tile. Perspective effects are evident in the frames but were ignored; some size estimates were therefore subject to errors when the selected tile was not at the same distance from the camera as the walker. For a further check, walking speed values derived from the gait parameters could be compared with average treadmill speeds given in the database documentation. Because walker progress was simulated by image displacement, body height and stride were not affected by perspective.

8.2 Results

Analyses were compared for one subject viewed from several angles. Height and stride features were consistent between directions, but limp and running tendencies were not detectable among random fluctuation.

9. Tests on conventional sequences

9.1 Data collection

Long video sequences were recorded in an indoor work area where many people pass to and fro, crossing the area in 12 steps or fewer. Shorter sequences, each showing a single passage, were extracted and the walkers identified by inspection. The sequences were converted to silhouettes by the method of Section 6.1, but it was found that the walkers could not be reliably distinguished from their backgrounds. Shadows, highlights and similarities of clothing and background all contributed to the problem. (In some sequences, feet passed behind foreground objects, creating false footprints at the tops of the objects. This fault is beyond the reach of background subtraction improvements.)

The methods of Sections 6.2 and 6.3 were then tried, and much improvement was gained, but the silhouettes were still not reliable enough for the detection of heads and feet. To test the gait parameters of Section 7.3 as a personal identification method, or as a means of deciding whether two unknown walkers are the same person or not, a manual approach was tried.

A simple software package was written to allow a human operator to step through a video sequence, find a walker and record in a file when and where each footstep started and ended, and where the top of the head was at each mid-stance or double-stance time. This "labelling" procedure was used on 204 sequences featuring 19 persons. It was often possible to judge where a foot landed even when it was hidden by a foreground object. The files produced were then used to generate synthetic silhouettes, one for each video frame, as shown in Figure 5. (These silhouettes were not intended to be realistic, merely to allow existing software to detect the relevant footprints and head positions. They proved adequate for that purpose.)

The nine gait parameters of Section 7.3 were estimated and the walker identified for each sequence. Some walkers appeared much more frequently than others.

9.2 Confusion analysis

The nine gait parameters have different physical dimensions and variabilities. Without doubt there is also a relationship between stride, cadence and speed, and another between minimum and maximum head heights. To allow variation between walkers to be compared with typical variation between sequences of one walker, the following procedure was followed:

1. The mean parameter vector was found for each walker and subtracted from all vectors for that walker to give residual vectors.
2. The covariance matrix for all the residual vectors was calculated, giving a measure of parameter variation and correlation averaged over all walkers.
3. A linear transformation was applied to all parameter vectors so that the covariance matrix became a unit matrix, that is, the residual components became equally variant and uncorrelated. (This is the principal component approach applied to the residual vectors.)
4. The Euclidean distance in the transformed vector space was used to compare vectors.

In the ideal case where all walkers are the same person and the parameters follow a joint normal distribution, this procedure would lead to a squared Euclidean distance proportional to a chi-squared variate. In the present case, the distance can be compared with a threshold to decide whether two sequences are of the same walker or of two different ones. The procedure can be considered as target detection, where "target" means "difference in identity", "hit" means "detecting a real difference", and "false alarm" means "deciding that two sequences of one walker are of different walkers".

Since the identities of all walkers are known, hits and false alarms can be detected for the test data for each threshold value. All combinations of two sequences are tested. The false alarm rate R_{FA} is the fraction of pairs of sequences of the same walker for which the distance is above the threshold, while the detection rate R_D is the fraction of pairs of sequences of different walkers for which the distance is above the threshold. Both rates increase as the threshold is reduced, and they may be plotted against the threshold, or against each other as a Receiver Operating Characteristic (ROC). Figures 6 and 7 show examples of both plots.

The plots can be used in several ways. If the gait parameters are useless for distinguishing walkers, the detection rates should be the same, with $R_D = R_{FA}$ for the ROC. If they separate them well, there should be a threshold value for which R_D is close to 1 and R_{FA} is close to 0. The ROC can show that such a value exists, while the other plot shows what value should be used for similar data. Further, if some of the gait parameters are removed from the data, or some of the principal components are removed because they are weak and represent the differences of correlated parameters, then the plots show how well the gait analysis performs without them.

Plots were prepared for many subsets of the gait parameters or their principal components and compared. Of particular interest was the use of head heights alone, for without the other parameters the analysis is greatly simplified. The effect of having some walkers much more often than others was investigated by removing some records of the more frequent walkers from the analysis to partly equalise the frequencies.

9.3 Results

The use of all nine gait parameters, shown in Figures 6 and 7, with a distance threshold near 5, allowed a pair of sequences in the collected data to be classified as "same walker" or "different walker" with error rates about 22% at the equal error point $1 - R_D = R_{FA}$. Equalising the frequencies of walkers made little difference to this result.

Using the head heights alone gave equal error rates close to 22% with a threshold of 2. Using just the minimum height increased the rates to about 28%. Using all gait parameters except the heights gave error rates about 35% at the equal error point with threshold 4, as shown in Figure 8.

Removing the directions of the weakest principal components of the residual vectors made little difference to the error rates until three out of nine had been removed. After that the rates began to increase. With four components retained, the results were much worse than those for the two components derived from head heights alone.

Scatter diagrams of pairs of gait parameters showed that some walkers were distinctive in some parameter values, and that those parameters might still be useful for separation of identities in a minority of cases.

10. Discussion

Clearly, for the selection of walkers in the test sequences, the two head heights are doing most of the work of distinguishing walkers. The extra work of extracting other gait parameters from video sequences, while it may sometimes improve the results, did not prove its worth in these experiments.

Eliminating weaker principal components of the residual vectors from the data did not make a reliable improvement. The fact that four components gave worse results than two head heights suggests that this approach may remove parameters with low errors of measurement, when they should be retained for that same reason.

Ben Abdelkader *et al* (2002) used a different evaluation technique for their results, finding that the correct walker out of 41 walkers appeared in the best 12 matches about 90% of the time, when they used the two head heights, cadence and stride. The evaluation appears more relevant to finding a person in a "gallery" of known persons.

11. Conclusions

A means of extracting measures that describe the gait of a person in a video sequence, regardless of the direction of walking, has been identified. It requires that the sequence be obtained with a fixed camera looking down at a shallow angle from above head height, and reduced to a sequence of silhouettes.

The ability of these measures to distinguish persons is mainly due to the estimation of their heights. The other measures are weak without the heights and make little difference when the heights are already in use, for the test sequences used. If further evidence of the usefulness of other gait parameters is required, a much larger set of test sequences will be needed. Ultimately, reliable results will depend on reliable, automatic silhouette extraction, which remains a difficult problem in some surveillance environments.

12. References

Akihara Y, Sagawa R, Mukaigawa Y, Echigo T & Yagi Y (2006), "Gait Recognition Using a View Transformation Model in the Frequency Domain", Proceedings European Conference on Computer Vision 2006, Lecture Notes on Computer Science 3953, Springer-Verlag, 151-163.

Bashir K, Xiang T & Gong S (2008), "Feature Selection on Gait Energy Image for Human Identification", Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing 2008, 985-988

Ben Abdelkader C, Cutler R & Davis L (2002), "View-invariant Estimation of Height and Stride for Gait Recognition", Lecture Notes on Computer Science 2359, Springer-Verlag.

Bhanu B (2007), "Human Recognition at a Distance", Proceedings SPIE 6786, article 678602, 1-6.

Bobick A F & Johnson A Y (2001) "Gait Recognition Using Static, Activity-Specific Parameters", Proceedings IEEE Conference on Computer Vision and Pattern Recognition 2001, 1, 423-430.

Bouchrika I & Nixon M S (2007), "Model-based Feature Extraction for Gait Analysis and Recognition", Proceedings MIRAGE 2007, Lecture Notes on Computer Science 4418, Springer-Verlag, 150-160.

Brubaker M A, Fleet D J & Hertzmann A (2007), "Physics-based Person Tracking Using Simplified Lower-body Dynamics", Proceedings IEEE Conference on Computer Vision and Pattern Recognition 2007, 1-8.

Cheng M-H, Ho M-F & Huang C-L (2008), "Gait Analysis for Human Identification Through Manifold Learning and HMM", Pattern Recognition 41, 2541-2553.

Collins R T, Gross R & Shi J (2002) "Silhouette-based Human Identification from Body Shape and Gait", Proceedings IEEE International Conference on Automatic Face and Gesture Recognition 2002, 366-371.

Gavrila D.M. (1999), "The Visual Analysis of Human Movement: A Survey", Computer Vision and Image Understanding, 73(1), 82-98.

Han J, Bhanu B & Roy-Chowdhury A K (2005), "A Study on View-insensitive Gait Recognition", Proceedings IEEE International Conference on Image Processing 2005, 3, 297-300.

He Q & Debrunner C (2000) "Individual Recognition from Periodic Activity Using Hidden Markov Models", Proceedings IEEE Workshop on Human Motion 2000, 47-52.

Huang P S, Harris C J & Nixon M S (1998), "Canonical Space Representation for Recognising Humans by Gait and Face", Proceedings IEEE South-West Symposium on Image Analysis & Interpretation, 180-185.

Kale A, Chowdhury A K R & Chellappa R (2003), "Towards a View Invariant Gait Recognition Algorithm", Proceedings IEEE Conference on Advanced Video and Signal Based Surveillance 2003, 143-150.

Kaziska D & Srivastava A (2006), "Cyclostationary Processes on Shape Spaces for Gait-based Recognition", Proceedings European Conference on Computer Vision, 2006, Lecture Notes on Computer Science 3953, Springer-Verlag, 442-453.

Knossow D, Ronfard R, Horaud R & Devernay F (2006) "Tracking with the Kinematics of Extremal Contours", Proceedings of ACCV 2006, Lecture Notes on Computer Science 3851, Springer-Verlag, 664-673.

Kuno Y, Watanabe T, Shimosakoda Y & Nakagawa S (1996), "Automated Detection of Human for Visual Surveillance System", Proceedings IEEE International Conference on Pattern Recognition 1996, 985-989.

Lee S., Liu Y & Collins R (2007), "Shape Variation-based Frieze Pattern for Robust Gait Recognition", Proceedings IEEE Conference on Computer Vision and Pattern Recognition 2007, 1-8.

Li H-G, Shi C-P & Li X-G (2005), "LLE based Gait Recognition", Proceedings IEEE International Conference on Machine Learning and Cybernetics 2005, 4516-4521

Moeslund T B & Granum E (2001), "A Survey of Computer Vision-Based Human Motion Capture", Computer Vision and Image Understanding, 8, 231-268.

Nixon M.S., Tan T.N. & Chellappa R. (2006), "Human Identification Based on Gait", Springer, New York.

Rao C & Shah M (2001), "View-Invariance in Action Recognition", Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2001, 2, 316-322.

Redding N R, Ohmer J F, Kelly J & Cooke T (2008), “Cross-Matching via Feature Matching for Camera Handover with Non-Overlapping Fields of View”, Proceedings Tenth International Conference on Digital Image Computing Techniques and Applications (DICTA 2008)

Shan C, Gong S & McOwan P W (2007), "Learning Gender from Human Gaits and Faces" Proceedings IEEE Conference on Advanced Video and Signal Based Surveillance 2007, 505-510.

Spencer N & Carter J (2005), "Towards Pose Invariant Gait Reconstruction", Proceedings IEEE International Conference on Image Processing 2005, 3, 261-264.

Tan D, Huang K, Yu S & Tan T (2007), *Recognising Night Walkers Based on One Pseudoshape Representation of Gait", Proceedings IEEE Conference on Computer Vision and Pattern Recognition 2007, 1-8.

Tomasi C & Kanade T (1992), “Shape and motion from image streams under orthography: A factorization method”, International Journal of Computer Vision, 9(2), 137-154.

Tresarden P A & Reid I D (2008), "Camera Calibration from Human Motion", Image and Vision Computing, 26, 851-862

Wang L., Hu W. & Tan T. (2003), "Recent Developments in Human Motion Analysis", Pattern Recognition 36, 585-601.

Wang L, Ning H, Hu W & Tan T (2002) "Gait Recognition Based on Procrustes Shape Analysis", Proceedings IEEE International Conference on Image Processing 2002, III, 433-436.

Yilmaz A & Shah M (2008), "A Differential Geometric Approach to Representing the Human Actions", Computer Vision & Human Understanding, 109, 335-351.

Zhou X & Bhanu B (2008), "Feature Fusion of Side Face and Gait for Video-based Human Identification", Pattern Recognition 41, 778-795.

13. Acknowledgments

Drs D. Kamenetsky and T. Scoleri assisted by labelling of some of the video sequences of walkers.

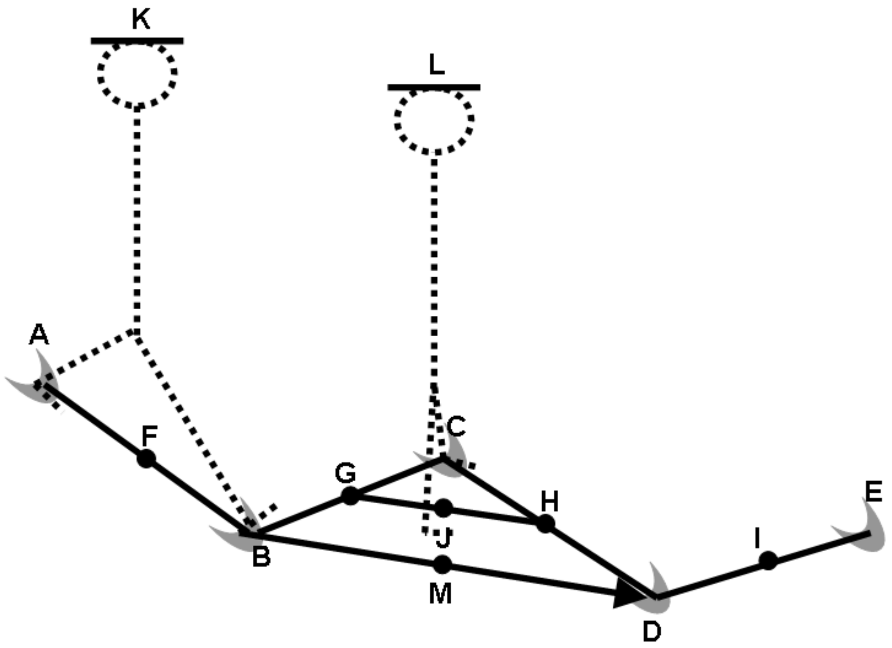

Figure 1. *Gait analysis of a series of footprints. A-E are footprint positions. F-I are midway between consecutive footprints. J is midway between G and H.* $\overrightarrow{GH} = \frac{1}{2}\overrightarrow{BD}$ *is the stride vector and the corresponding time difference is the stride time. F marks a body position when the head is lowest, at K. J marks a body position when the head is highest, at L.* $\overrightarrow{JC}$ *is to the left from* $\overrightarrow{GH}$*, showing that C is a left footprint.* M *is midway between B and D*

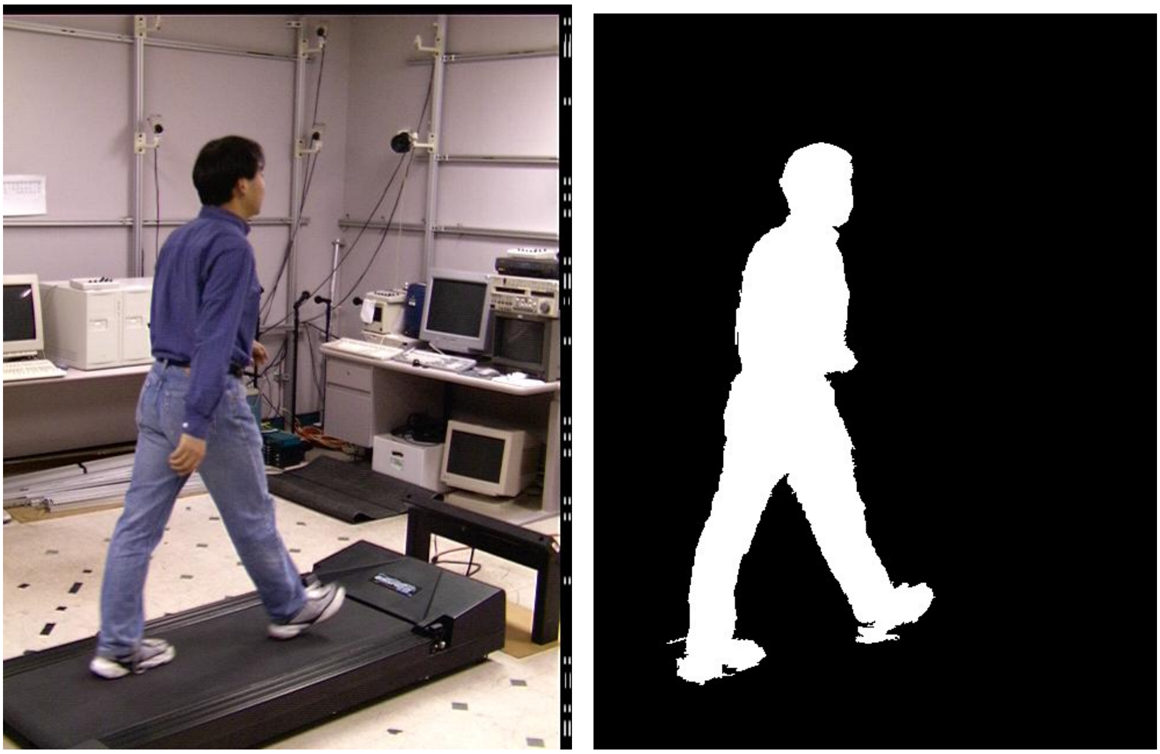

Figure 2: A frame from the MoBo database and the silhouette extracted by the method of Section 6.1.

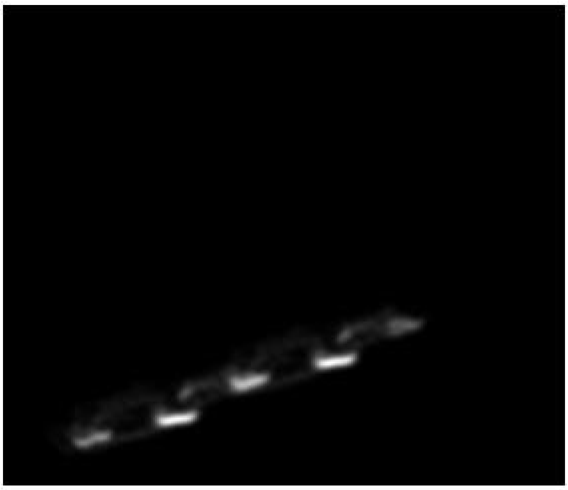

Figure 3: Sums of edge values from the same treadmill sequence as Figure 2, modified to simulate ground walking.

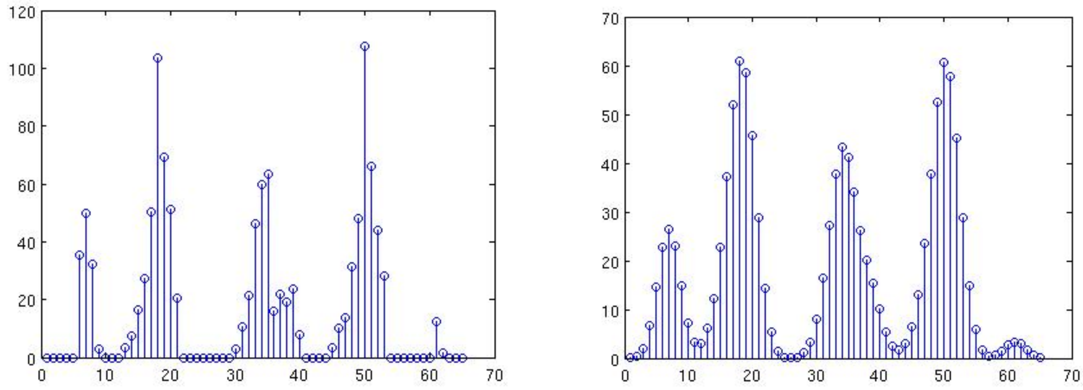

Figure 4: Histogram of mean times of footprint pixels, before and after smoothing.

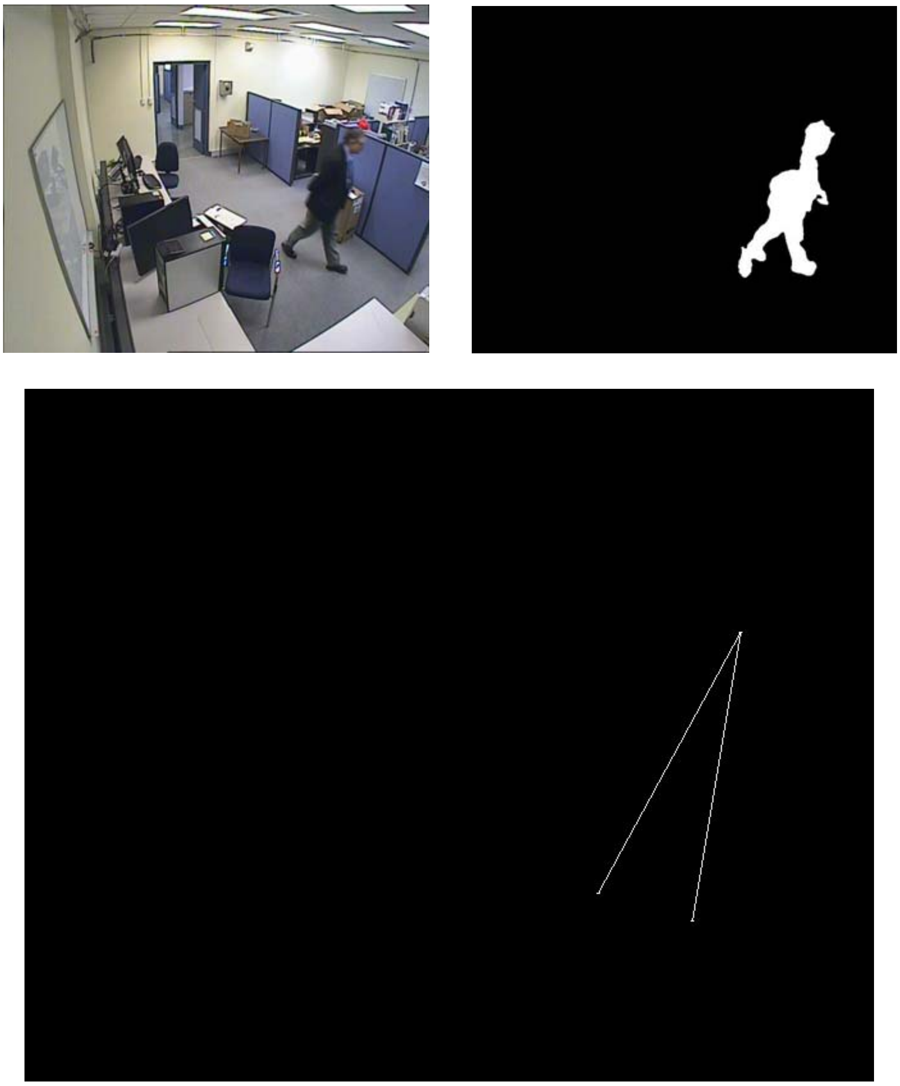

Figure 5: Synthetic silhouette generation.. Top left: walker and scene. Top right: best real silhouette. Bottom: synthetic silhouette.

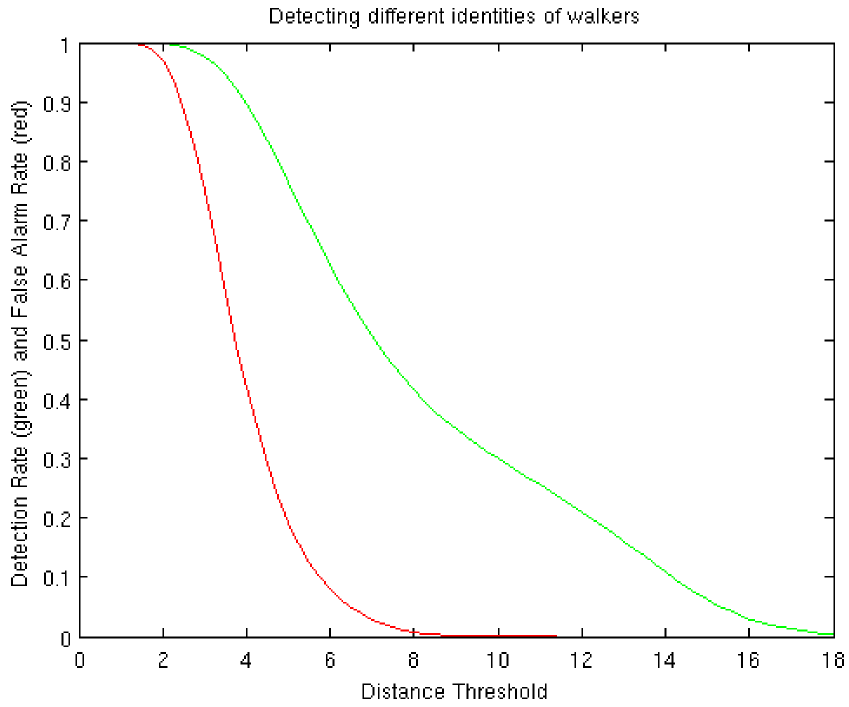

Figure 6: Detection and false alarm rates for detecting different identities, plotted against distance threshold when all gait parameters used

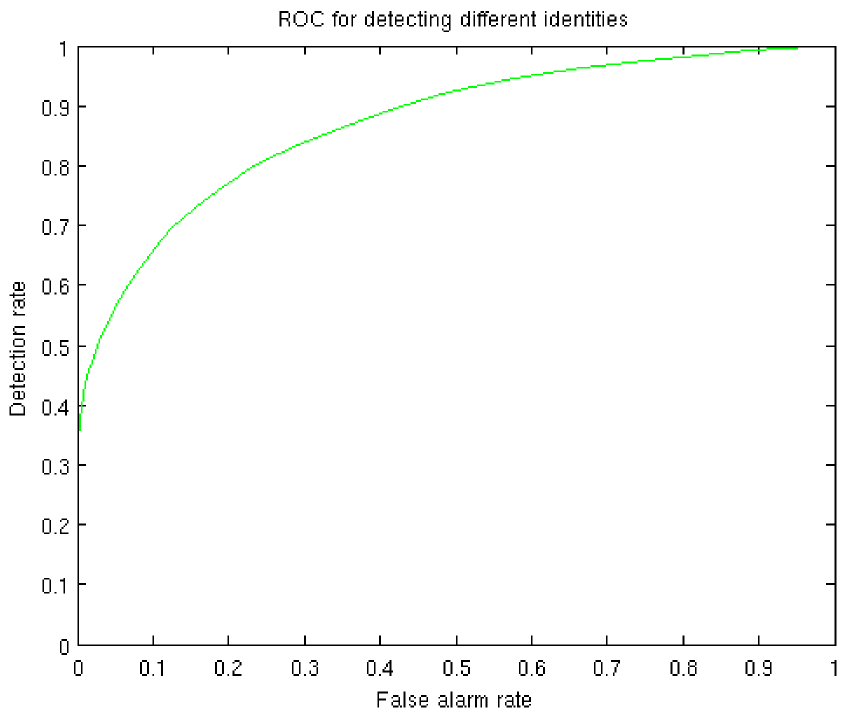

Figure 7: ROC for detecting different identities when all gait parameters used..

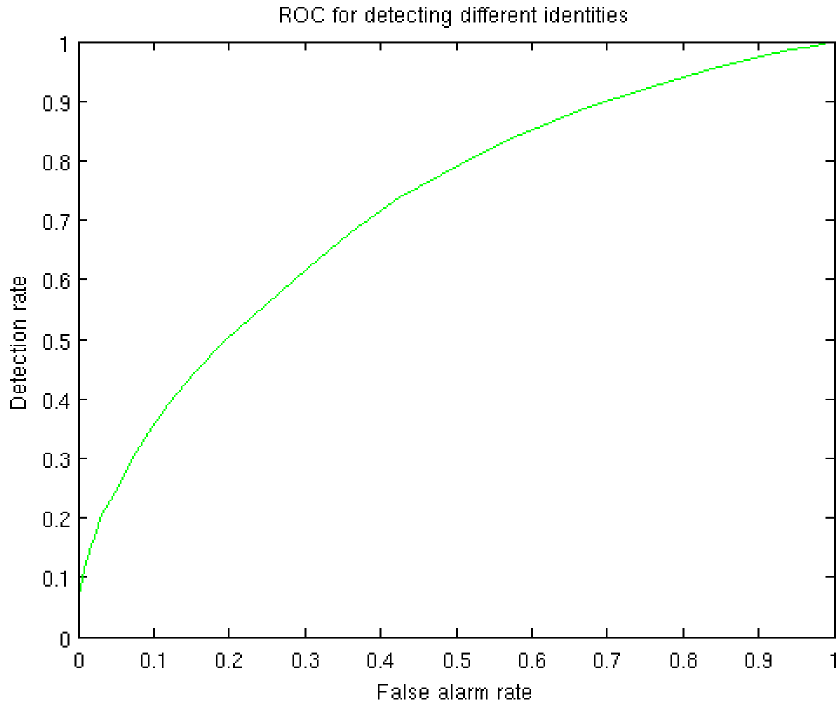

Figure 8: ROC for detecting different identities when gait parameters other than heights used..

Page classification: UNCLASSIFIED

<table>
<tr><td colspan="3">DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA</td><td colspan="3">
1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)</td></tr>
<tr><td colspan="3">2. TITLE

A Feasibility Study of View-independent Gait Identification</td><td colspan="3">3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION)

Document (U)
Title (U)
Abstract (U)</td></tr>
<tr><td colspan="3">4. AUTHOR(S)

Robert Whatmough</td><td colspan="3">5. CORPORATE AUTHOR

DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh South Australia 5111 Australia</td></tr>
<tr><td>6a. DSTO NUMBER
DSTO-TR-2684</td><td colspan="2">6b. AR NUMBER
AR-015-265</td><td colspan="2">6c. TYPE OF REPORT
Technical Report</td><td>7. DOCUMENT DATE
March 2012</td></tr>
<tr><td>8. FILE NUMBER
2010/1173996/1</td><td>9. TASK NUMBER
07/114</td><td>10. TASK SPONSOR
DCDS(IWS)</td><td colspan="2">11. NO. OF PAGES
26</td><td>12. NO. OF REFERENCES
32</td></tr>
<tr><td colspan="3">13. DSTO Publications Repository

http://dspace.dsto.defence.gov.au/dspace/DSTO-TR-2684</td><td colspan="3">14. RELEASE AUTHORITY

Chief, Intelligence, Surveillance and Reconnaissance Division</td></tr>
<tr><td colspan="6">15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111</td></tr>
<tr><td colspan="6">16. DELIBERATE ANNOUNCEMENT

No Limitations</td></tr>
<tr><td colspan="6">17. CITATION IN OTHER DOCUMENTS Yes</td></tr>
<tr><td colspan="6">18. DSTO RESEARCH LIBRARY THESAURUS

Biometric identification; Feature extraction; Image sequences, Image classification; Visual surveillance</td></tr>
<tr><td colspan="6">19. ABSTRACT
This report reviews work on the identification of people in video recordings by their gait, or walking style. It considers the adjustments needed when the subjects walk other than straight across the image and may not offer the convenient side (fronto-parallel) view. It describes a study of this case using video sequences of walking people, and concludes that the available measurements of gait, apart from head height, are of limited use.</td></tr>
</table>

Page classification: UNCLASSIFIED